



SCI 论文作者甄别软件设计及应用*

于 健¹ 吴 霞² 赵春梅²

¹(中国科学院国家科学图书馆 北京 100190)

²(中国科学院高能物理研究所 北京 100049)

摘要:【目的】结合机构产出 SCI 论文统计需求,设计一款自动甄别目标机构作者和实验室的软件。【应用背景】可辅助论文统计部门快速准确识别机构论文作者和实验室(部门),进而获得机构作者和实验室的论文产出分布情况。【方法】从技术上实现综合利用相同研究单元内作者合作较多的科研特点、自定义作者唯一关键词或合作者字段以及 SCI 数据库作者相关字段的文本特征来甄别目标机构作者。【结果】允许用户通过目标机构人员名单维护来实现 SCI 论文作者甄别的自动化和高准确度。【结论】有效解决 SCI 论文中文作者因拼音写法多样且易重名而造成作者相关论文数据难以准确统计的问题,其设计思路也适用于 EI 及其他数据库论文作者甄别。

关键词: 论文统计 作者甄别 SCI 软件设计

分类号: G356

1 引言

近年来,高校和科研机构愈加重视机构产出 SCI 学术论文的统计管理和分析利用^[1-3],也有越来越多的机构为此参与建设了机构知识库等学术资产管理系统^[4]。目前机构产出 SCI 论文统计服务通常有两种形式:先由作者自行提交论文,后由科研论文统计部门核查并汇总;先由科研论文统计部门搜集整理论文,后由作者确认后汇总。据一份调查结果显示,多数科研人员都有找不到科研成果的经历^[4],他们为了节省时间和精力往往提交不够标准的成果数据,因此会希望科研统计部门辅助统计其成果数据,而他们只需从中复核挑选自己的论文。但是,科研论文统计部门在进行 SCI 论文统计时常会面临一个棘手的问题:如何甄别本机构的 SCI 论文作者?尤其中文作者拼音写法具有多样性,易出现重名和引发混淆,从而导致即便机构管理系统里存储了大批 SCI 论文数据却无法快速

准确得到某作者或实验室(部门)的产出情况^[5]。本文所述 SCI 论文作者甄别软件可通过综合利用目标机构人员名单中的作者不同拼音写法、合作者、自定义唯一关键词和合作者字段来实现目标机构作者的准确自动识别。甄别后的 SCI 论文数据表上传到机构知识库或者直接进行分析后,可实现准确、快速统计目标机构作者和实验室 SCI 论文的产出情况^[6]。

2 软件需求

目前,许多机构论文管理系统的作者识别需要在很大程度上依赖人工逐一判断确认^[7,8],其中有的利用机构作者的各种拼音写法来辅助人工进行确认,例如中国科学院 IR 机构知识库系统,在一定程度上提高了论文确认工作效率。此外,目前已有很多关于文献重名作者处理方法的研究^[9,10],研究领域涉及文献检索、文献数据库等,研究对象是海量文献的重名作者处理,

收稿日期: 2013-10-22

收修改稿日期: 2014-01-22

*本文系中国科学院研究所情报分析可持续服务能力建设子项目“中科院高能所情报分析可持续服务能力建设”(项目编号: 院 1105)和中国科学院国家科学图书馆青年人才领域前沿项目“学科化知识服务辅助工具优化设计”(项目编号: Q1209)的研究成果之一。

主要是基于文献的合作者关系以及文献文本属性特征(包括题名、来源出版物、主题词等)来处理重名问题,有的也达到了较高的准确率,但是仍无法满足知识管理领域的机构这类特定对象的论文统计准确率和灵活性的需求。

本软件的特点主要在于专门针对机构论文统计的需求,在从微观角度深入挖掘和利用 SCI 数据库文献特征(包括作者全简称写法和地址字段)的基础上,充分利用研究机构内同研究单元(同课题组或实验室)内作者合作较多的科研规律,并且通过机构内重名作者的自定义唯一关键词或合作者来实现目标机构作者的自动和高准确率识别。

3 软件功能模块设计

本软件设计思路是基于机构人员名单(含所属课题组和实验室)来甄别 SCI 论文作者,主体结构由以下三个模块构成。

3.1 机构人员名单维护模块

根据实际需求,本模块考虑了团体作者和非团体作者两种不同情形,并相应设置了两机构人员名单所需字段,如图 1 所示。

非团体作者含 5 个字段,分别是汉字写法、拼音全称一(姓名所有单字以空格隔开)、拼音全称二(仅姓和名以空格隔开)、拼音简写一(姓全拼+空格+名所有单字首

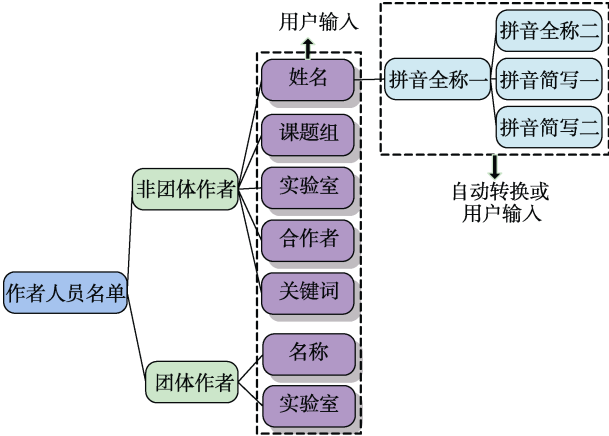


图 1 机构人员名单字段

字母)和拼音简写二(姓全拼+空格+名首字的首字母)。

非团体作者包含了两个附加信息字段,即自定义唯一合作者和自定义唯一关键词字段,帮助用户应对软件甄别结束后甄别结果中仍然存在重名作者的情况。此时,用户需要人工甄别重名作者,在甄别完成后,可据此甄别经验,在机构人员名单中自定义输入用以区分该重名作者的合作者或关键词。

3.2 目标机构作者提取模块

本模块将首先识别 SCI 论文的作者是否为团体作者,即 SCI 论文数据表中的字段 CA 是否不空。若是团体作者,则直接进入下一个作者甄别模块;若是非团体作者,则首先提取目标机构全部作者并进入下一模块。

表 1 SCI 论文数据表非团体作者相关字段

No	AU	AF	C1
1	Wang, YF; Qian, S; Zhao, T; Tian, JS; Li, HR; Cao, J; Xu, XY; Wang, XY; Liu, SD; Liu, HL; Liu, SL; Liu, DL; Heng, YK; Cao, XB; Jun, ST	Wang, Yifang; Qian, Sen; Zhao, T.; Tian, Jinshou; Li, Huirui; Cao, Jun; Xu, Xiangyan; Wang, Xiaoyun; Liu, Shudong; Liu, Hulin; Liu, Shulin; Liu, Delin; Heng, Yuekun; Cao, Xibin; Jun Shentu	[Wang, Yifang; Qian, Sen; Zhao, T.; Cao, Jun; Liu, Shudong; Liu, Shulin; Heng, Yuekun] Chinese Acad Sci, Inst High Energy Phys, Beijing 100049, Peoples R China; [Tian, Jinshou; Xu, Xiangyan; Liu, Hulin] Chinese Acad Sci, Xian Inst Opt & Mech, Xian 710119, Peoples R China; ...
2	Wang, P; Leinweber, DB; Thomas, AW; Young, RD	Wang, P.; Leinweber, D. B.; Thomas, A. W.; Young, R. D.	[Wang, P.] Chinese Acad Sci, Inst High Energy Phys, Beijing 100049, Peoples R China; [Wang, P.] Chinese Acad Sci, Theoret Phys Ctr Sci Facil, Beijing 100049, Peoples R China; [Leinweber, D. B.; Thomas, A. W.; Young, R. D.] Univ Adelaide, Sch Chem & Phys, Special Res Ctr Subatom Struct Matter CSSM, Adelaide, SA 5005, Australia; ...
3	Chen, C; Ge, C	Chen, C.; Ge, C.	[Chen, C.] Chinese Acad Sci, Natl Ctr Nanosci & Technol, Beijing, Peoples R China; [Ge, C.] Chinese Acad Sci, Inst High Energy Phys, Beijing, Peoples R China
4	Dong, YB	Dong Yu-Bing	Chinese Acad Sci, Inst High Energy Phys, Beijing 100049, Peoples R China

SCI论文数据表中字段C1的格式共有两种,如表1所示,即作者(写法同字段AF)+作者机构或者仅作者机构。当C1字段为作者+作者机构时,本模块将根据SCI论文数据表中的字段C1和用户从软件界面输入的目标机构名称的唯一标识词来识别并提取出作者。当C1字段只含作者机构时,表明所有作者同属于一个机构,本模块将根据SCI论文数据表中的字段AF和目标机构唯一标识词提取出作者。

3.3 作者甄别执行模块

对于团体作者论文,本模块将根据机构人员名单中的团体作者信息字段进行甄别。对于非团体作者论文,

本模块根据机构人员名单中的人员信息字段来甄别。在完成一篇论文的所有作者甄别后,本模块将进行该论文所属实验室的赋值,赋值原则为首先选取第一作者的实验室为该论文所属实验室,若第一作者实验室不存在,则选取作者最多的实验室为该论文所属实验室。

3.4 软件实现流程

本软件是在VB 6.0环境下开发的,完整流程从调用机构人员名单 Excel 文件对象和SCI论文数据表 Excel 对象开始,到甄别完成后将甄别作者姓名和所属实验室值写入SCI论文数据表 Excel 文件对象结束,如图2所示:

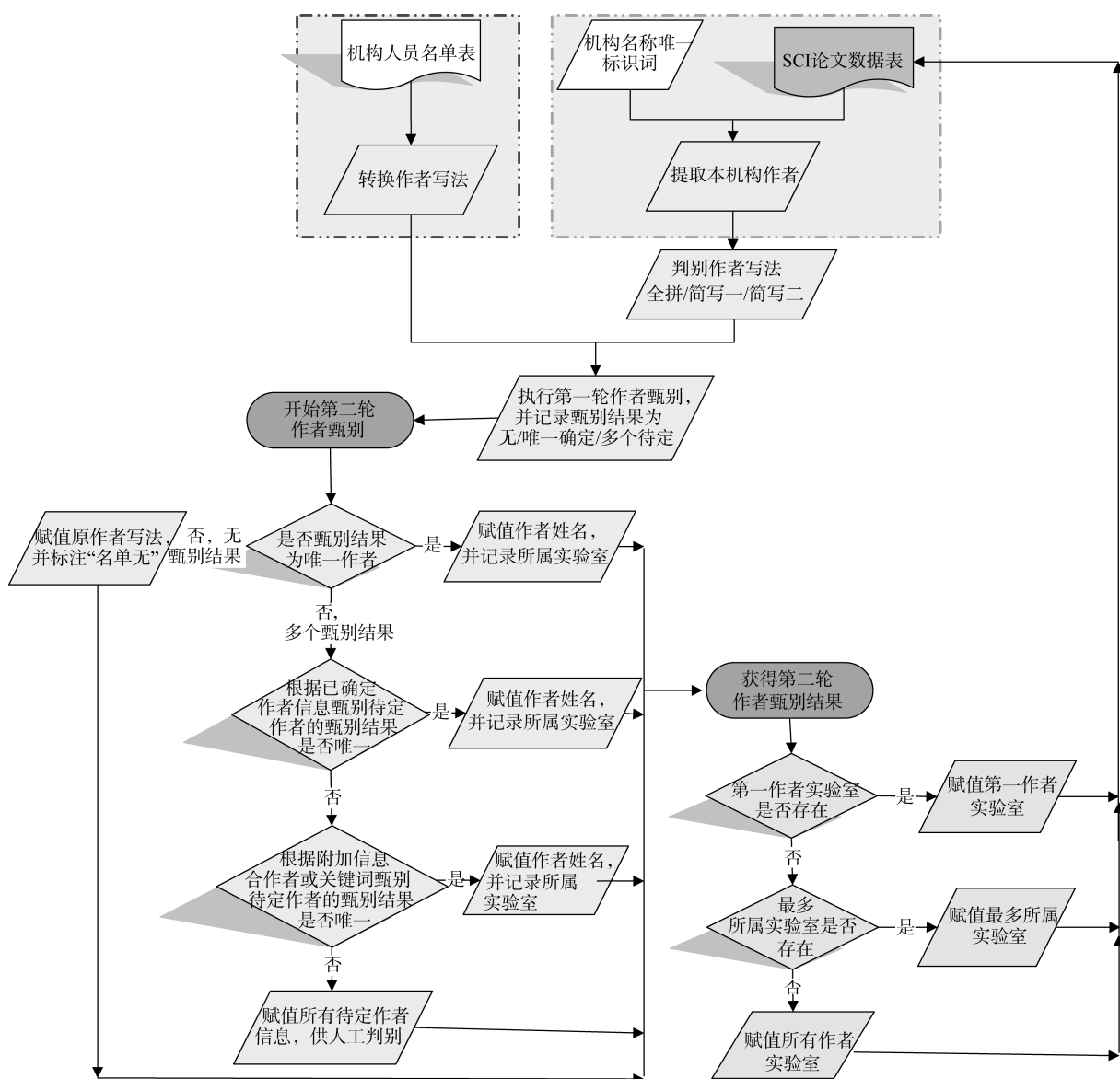


图2 软件流程

4 软件技术实现的关键问题

4.1 机构人员名单中作者全简称写法的自动转换

软件将用户输入人员姓名自动转换为拼音全称^[11]，继而转换为 SCI 数据库中的拼音全称和拼音简写，以作者“钱学森”为例，作者姓名的转换过程为：“钱学森—qian xue sen—qian xuesen—qian xs—qian x”；英文姓名全称的自动转换过程相对较为简单，以“albert einstein”为例，其转换过程为：“albert einstein—albert einstein—albert einstein—albert e—albert e”。

为了解决多音姓氏问题，本软件内置了常见多音姓氏的自动转换，包括黑/仇/区/朴/折/单/查/解/繁/缪/员/种/笪/万俟/单于/尉迟等。如果用户的机构人员名单中存在其他多音字，则需要人工输入含多音字作者的拼音写法。

4.2 SCI 论文目标机构作者的正确提取

本软件通过 SCI 论文表中 C1 字段来识别目标机构作者，因此正确提取作者的关键是解决目标机构写法的输入问题。

为了方便用户使用，软件识别目标机构写法的设计思路与在 SCI 数据库中检索目标机构论文的方法相似。用户只需根据目标机构的地址在 SCI 数据库中的写法(包括机构名称、城市、国家或邮编)组合成机构唯一标识词进行输入即可。例如，中国科学院高能物理所(简称“中科院高能所”)在 SCI 数据库中的通常写法为“Chinese Acad Sci, Inst High Energy Phys, Beijing 100049, Peoples R China”。根据工作经验，除了中科院+高能所的标准写法以外，高能所+北京或高能所+邮编(100049 或 100039)均大致可唯一确定为中科院高能所，只需排除成都电子科技大学高能电子研究所、北京大学高能物理研究中心、清华大学高能物理研究中心即可；此外，中科院高能所内的某些实验室，例如同步辐射实验室+北京或同步辐射实验室+邮编(100049 或 100039)也可唯一确定为中科院高能所，因此在 SCI 数据库中检索高能所论文时，通常会使用检索式“(“Inst High Energy Phys” OR ihep OR “Beijing Synchrotron Radiat” OR BSRF) SAME (beijing OR china OR chinese OR cas OR 100039 OR 100049) NOT (Chengdu OR Peking OR “Univ Beijing” OR Tsinghua OR Qinghua)”来进行检索。同理，本软件允许用户在

软件使用界面做类似的目标机构输入，便可准确识别目标机构并提取作者。

4.3 重名作者的甄别问题

为了解决作者重名写法的问题，软件对所提取作者进行两轮甄别，如图 3 所示：

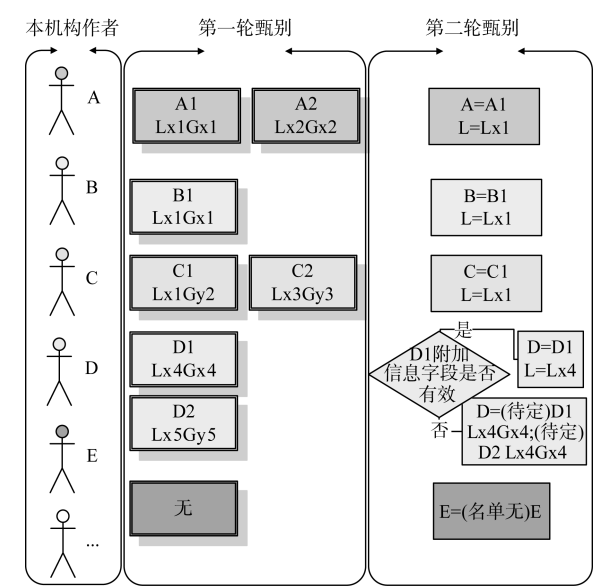


图 3 重名作者甄别过程示例模型
(注：A 表示作者；L 表示实验室(部门)；G 表示课题组)

第一轮首先确定提取作者的拼音全称或简写，并据此进入机构人员名单中进行甄别，记录下每位作者的所有可能作者信息，包括姓名、课题组和实验室。

SCI 论文数据表中字段 AU 是 SCI 数据库加工人员规范化后的作者拼音简写，字段 AF 是 SCI 论文原文中的作者写法，可能是拼音全称或拼音简写。因此，软件可通过对比 AF 和 AU 两个字段来判断是拼音全称(如表 1 中论文 1 所示)或是简写。若是简写，则继续判断是拼音简写一(如表 1 中论文 2 所示)还是拼音简写二(如表 1 中论文 3 所示)。

第二轮甄别分三种情况处理：若第一轮未甄别到任何作者信息，则为该作者最终甄别结果赋值 SCI 论文中原作者写法，并标注“名单无”；若第一轮甄别到唯一作者，直接赋值该作者姓名；若第一轮甄别到多个作者，则首先根据其合作者中已确定作者的课题组或实验室来进行甄别，若是仍无法确定，则根据附加信息字段自定义唯一合作者或关键词来进行甄别，最后若仍无法确定，则为该作者最终甄别结果赋值所有

可能的作者信息,并标注“待定”。

5 应用效果评估

笔者使用该软件来测试甄别中国科学院高能物理研究所 2008 年~2012 年发表 SCI 论文的作者和所属实验室。

首先使用软件包自带的 Excel 文件机构人员名单模板来制作一份目标机构人员名单,如图 4 所示。软件所需的另外一个 Excel 文件 SCI 论文数据表可以从 SCI 数据库下载获取,在 SCI 数据库中保存文件格式为 Tab-delimited(win)并导入 Excel 文件即可。若已保存文件格式为 Plain Text,则需要使用 SCI 转换工具转换为本软件所需的 Excel 文件^[12]。

姓名	实验室	自定义唯一合作者或自定义唯一关键词	自定义唯一合作者或自定义唯一关键词
1. 周俊杰	加速器研究中心		ALFPI Collaborator: 实验物理中心
2. 黄永	加速器研究中心		AMS Q1 Collaborator: 粒子天体物理研究中心
3. 沈 翥	加速器研究中心		AMS Collaboration: 粒子天体物理研究中心
4. 杨兴旺	加速器研究中心		ARGO-YBJ Collaborator: 粒子天体物理研究中心
5. 刘金家	加速器研究中心		ASY Collaborator: 粒子天体物理研究中心
6. 何大勇	加速器研究中心		ATP Collaborator: 加速器中心
7. 廖 军	加速器研究中心		ATLAS Collaborator: 实验物理中心
8. 周国伟	加速器研究中心		BABAR Collaborator: 实验物理中心
9. 刘 俊	加速器研究中心		Belle Collaborator: 实验物理中心
10. 廖振华	加速器研究中心		BES Collaborator: 实验物理中心
11. 吕 捷	加速器研究中心		BESIII Collaborator: 实验物理中心
12. 马新刚	加速器研究中心		CLQCD Collaborator: 理论物理研究室
13. 肖欣工	加速器研究中心		CMS Collaborator: 粒子天体物理研究中心
14. 赵永利	加速器研究中心		CMS Collaborator: 粒子天体物理研究中心
15. 刘晋浩	加速器研究中心		DO Collaborator: 实验物理中心
16. 陈 忠	加速器研究中心		Dura Bar: 实验物理中心
17. 陈敬如	加速器研究中心		DELPHI Collaborator: 实验物理中心
18. 王瑞蒙	加速器研究中心		HERA-B Collaborator: 实验物理中心
19. 张 祥	加速器研究中心		High Resolution: 粒子天体物理研究中心
20. 李小平	加速器研究中心		KamLAND Collaborator: 实验物理中心
21. 张翼	加速器研究中心		LDM Collaborator: 实验物理中心

图 4 机构人员名单

在软件界面输入中科院高能所名称的唯一标识词,如图 5 所示,并选择机构人员名单和 SCI 数据表 Excel 文件路径,单击按钮开始作者甄别,甄别完成以后的 SCI 数据表如图 6 所示。

从名单初始化到甄别完成测试数据表约需 1 分钟,图 7 显示了此次测试甄别作者的结果统计情况。全部论文的作者总人次约为 12 114,作者总人数约为 2 041,其中准确甄别到作者总人次约为 10 526,作者



图 5 软件界面

甄别作者	甄别实验室	A	B	C	D	E	F	G	H	I
1. 周俊杰	加速器研究中心	7	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.
2. 黄永	加速器研究中心	7	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.
3. 沈 翥	加速器研究中心	7	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.
4. 杨兴旺	加速器研究中心	7	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.
5. 刘金家	加速器研究中心	7	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.
6. 何大勇	加速器研究中心	7	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.
7. 廖 军	加速器研究中心	7	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.
8. 周国伟	加速器研究中心	7	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.
9. 刘 俊	加速器研究中心	7	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.
10. 廖振华	加速器研究中心	7	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.
11. 吕 捷	加速器研究中心	7	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.
12. 马新刚	加速器研究中心	7	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.
13. 肖欣工	加速器研究中心	7	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.
14. 赵永利	加速器研究中心	7	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.
15. 刘晋浩	加速器研究中心	7	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.
16. 陈 忠	加速器研究中心	7	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.
17. 陈敬如	加速器研究中心	7	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.
18. 王瑞蒙	加速器研究中心	7	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.
19. 张 祥	加速器研究中心	7	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.
20. 李小平	加速器研究中心	7	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.
21. 张翼	加速器研究中心	7	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.	Li, H.

图 6 SCI 论文表甄别结果

总人数约为 1 235, 标记(名单无)的作者总人次和总人数分别约为 1 190 和 684, 标记(待定)的作者总人次和总人数分别约为 398 和 122。从甄别结果可以看出, 不考虑用户工作经验所积累的 SCI 检索式写法完整度和 SCI 数据库地址字段标引错误的人为因素, 本软件的甄别准确性完全取决于机构人员名单的准确性, 只要机构人员名单中已有作者信息, 软件已经全部甄别到作者和实验室; 而对于名单中缺少的作者, 软件已标记(名单无), 供用户后续核实并在机构人员名单中添加, 这样软件今后也将能够甄别到该作者。对于软件无法唯一确定的重名作者, 软件已标记(待定), 用户可以根据人工甄别经验, 在机构人员名单的自定义唯一合作者或自定义唯一关键词字段中自行输入区分这些重名作者的合作者或关键词, 这样软件以后将能够据此来区分甄别。由此可见, 随着机构人员名单的完善, 软件的识别率将不断提高, 使机构论文统计人员通过日常经验积累脱离作者识别上的繁琐人工干预工作。

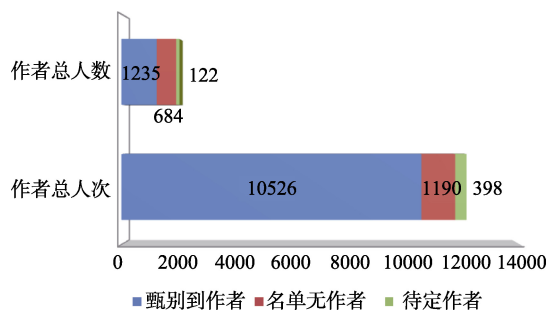


图 7 SCI 论文表的作者甄别结果统计

6 结 语

本文从实际需求出发, 开发了一款用于机构 SCI 论文统计服务的作者甄别软件, 以辅助提高科研论文管理效率和准确性。该软件除了实现甄别功能以外,

还可以通过人机交互和用户对机构人员名单信息的持续完善使软件甄别率不断提高。该软件在实现甄别功能的同时也具有清理机构论文数据的功能,排除用户目标机构名称唯一标识词输入不全的因素,提取不到目标机构作者信息的论文很可能不是目标机构所发表的论文。此外,该软件虽然是针对 SCI 论文统计服务设计的,但是也可直接应用于论文数据字段格式相同的 WOS 其他数据库(如 CPCI)的论文统计,其设计思路也同样适用于 EI 或其他数据库。

参考文献:

- [1] 包凌, 钟建文. 2001~2010 年清华大学水利水电工程系发表论文统计分析[J]. 情报探索, 2012(4): 49-53. (Bao Ling, Zhong Jianwen. Statistics and Analysis on Papers of Tsinghua University Published During 2001~2010[J]. Information Research, 2012(4): 49-53.)
- [2] 王玉芹. 基于 SCI-E 的中国农业科学院科研论文统计分析[J]. 农业图书情报学刊, 2012, 24(8): 54-59. (Wang Yuqin. Statistical Analysis of Publications of Chinese Academy of Agricultural Sciences Based on SCI-E[J]. Journal of Library and Information Sciences in Agriculture, 2012, 24(8): 54-59.)
- [3] 杨俊丽. 基于 SCI、SSCI 和 A&HCI 的河南大学论文统计[J]. 大学图书情报学刊, 2013, 31(1): 75-79. (Yang Junli. The Statistical Analysis of Essays Published by Henan University Based on SCI, SSCI, A&HCI[J]. Journal of Academic Library and Information Science, 2013, 31(1): 75-79.)
- [4] 曾苏, 马建霞, 汤天波, 等. 国内科研机构 and 高校机构知识库规划建设现状与问题研究[J]. 现代图书情报技术, 2009(1): 50-57. (Zeng Su, Ma Jianxia, Tang Tianbo, et al. Study on the Present Situation and Problems of Development of Institutional Repository in Chinese Scientific Research Institutions and Universities[J]. New Technology of Library and Information Service, 2009(1): 50-57.)
- [5] 丁娜. 2004 年至 2012 年北京交通大学 SCI 收录论文统计与分析[J]. 中国科教创新导刊, 2013(11): 1-2. (Ding Na. Statistics and Analysis on Papers of Beijing Jiaotong University Published During 2004~2012[J]. China Education Innovation Herald, 2013(11): 1-2.)
- [6] 刘巍, 祝忠明, 张旺强, 等. 机构知识库个性化知识资产统计服务的设计与实现研究[J]. 现代图书情报技术, 2012(4): 17-21. (Liu Wei, Zhu Zhongming, Zhang Wangqiang, et al. Development and Research of Personalization Knowledge Property Statistics Module for Institutional Repository[J]. New Technology of Library and Information Service, 2012(4): 17-21.)
- [7] 郭薇, 常健. 中国行政管理学研究作者成熟度的文献计量学分析[J]. 学海, 2008(4): 92-99. (Guo Wei, Chang Jian. Bibliometric Study on the Maturity of China's Administration Researchers[J]. Academia Bimestris, 2008(4): 92-99.)
- [8] 山东师范大学图书馆情报分析组. 山东师范大学科研论文全景分析[J]. 山东师范大学学报: 人文社会科学版, 2013, 58(1): 135-160. (SDNU Library Intelligence Analysis Group. General Analysis of Thesis of Shandong Normal University[J]. Journal of Shandong Teachers' University: Humanities and Social Sciences, 2013, 58(1): 135-160.)
- [9] 袁军鹏, 俞征鹿, 苏成, 等. 作者重名辨识研究进展[J]. 数字图书馆论坛, 2011(10): 60-65. (Yuan Junpeng, Yu Zhenglu, Su Cheng, et al. A Survey of Author Name Disambiguation[J]. Digital Library Forum, 2011(10): 60-65.)
- [10] 郭舒. 文献数据库中作者名消歧算法研究[J]. 现代图书情报技术, 2013(7-8): 69-74. (Guo Shu. Research on Author Name Disambiguation Algorithm in the Literature Database[J]. New Technology of Library and Information Service, 2013(7-8): 69-74.)
- [11] Excel Home. 汉字转拼音的完美解决方案[EB/OL]. [2011-09-23]. <http://club.excelhome.net/thread-229924-1-1.html>. (Excel Home. The Perfect Solution to Switching Chinese Characters to Pinyin [EB/OL]. [2011-09-23]. <http://club.excelhome.net/thread-229924-1-1.html>.)
- [12] 科学网. SCI 转换工具[EB/OL]. [2013-03-05]. <http://blog.sciencenet.cn/home.php?mod=space&uid=260374&do=blog&id=667402>. (ScienceNet.cn. Format Transformation Software for SCI Data[EB/OL]. [2013-03-05]. <http://blog.sciencenet.cn/home.php?mod=space&uid=260374&do=blog&id=667402>.)

作者贡献声明:

于健: 软件功能设计实现和论文撰写;
吴霞, 赵春梅: 参与软件功能需求分析和软件测试。
(通讯作者: 于健 E-mail: yuj@mail.las.ac.cn)

Design and Application of Author Discrimination Software on Papers Indexed by SCI

Yu Jian¹ Wu Xia² Zhao Chunmei²

¹(National Science Library, Chinese Academy of Sciences, Beijing 100190, China)

²(Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China)

Abstract: [Objective] The software to discriminate one scientific institute's authors of scientific papers is designed to meet demands of the statistics of papers indexed by SCI. [Context] It can be used to help the department of statistical analysis on papers in SCI to determine Chinese characters for the Chinese author name belong to their institute and its corresponding lab. [Methods] Author discrimination is implemented technically by the comprehensive utilization of one characteristics of scientific research that people from the same research units are more likely to co-author papers, custom unique keywords or co-authors and text features of author fields in SCI. [Results] Automation and high accuracy of author discrimination can be achieved based on maintenance of a personnel list of one scientific institute. [Conclusions] It effectively solves the duplication problem of Chinese names during the analysis of papers in SCI and its design ideas also apply to other databases such as EI and Inspec.

Keywords: Papers statistics Author discrimination SCI Software design

Innovative 和 Bibliotheca 宣布结成战略联盟，致力于整合图书馆技术解决方案

全球领先的图书馆服务平台提供商 Innovative，和全球领先的 RFID 和条码技术解决方案提供商 Bibliotheca，于近日宣布结成战略联盟，将加强两家公司产品的整合，为全球的图书馆用户创造附加价值。根据合作协议，Innovative 将有权直接转售 Bibliotheca 的 RFID 系统给 Innovative 全球的客户，包括美国、欧洲和澳大利亚。

两家公司同意共同合作，将双方的产品线互联互通，以促进技术创新。协议还允许 Innovative 使用“triple I”商标销售 Bibliotheca 的产品，并在双方公司的销售、售后支持和咨询部门之间建立更加正式的工作关系，这将有效地解决集成独立的 ILS (集成图书馆系统)和 RFID 技术时所带来的一系列常见问题。

“Bibliotheca 已经为图书馆建立了 RFID 解决方案的标准，我们很高兴能与他们合作，推进双方产品连通性。”Innovative CEO Kim Massana 说：“我们建立这种合作关系的终极目标是简化改善图书馆读者的体验，为我们的客户创建无缝的技术解决方案。”

“Innovative 是图书馆自动化解决方案的领导者，我们很自豪能通过他们为客户提供我们的整套产品。”Bibliotheca 总经理 Jim Hopwood 说：“我们很高兴能和 Innovative 一起连通我们的解决方案和他们的图书馆服务平台。”

Innovative 的旗舰图书馆服务平台——Sierra——是图书馆自动化史上扩散速度最快的一款产品。全球 1 500 多家图书馆已经部署 Sierra，巩固了 Innovative 在全球图书馆自动化市场的领先地位。

Bibliotheca 的产品线，以其备受赞誉的 liber8 触摸屏体验为主打，在全美已有 1 000 多家客户，全球已有 4 000 多家客户。Bibliotheca 的面向读者的特点，加之与 Sierra 平台的直接交互，这将为图书馆读者和员工创造无与伦比的用户体验。

(编译自: <http://www.iii.com/iii-bibliotheca>)

(本刊讯)